

Claims

1. A method for generating a statistic for a pattern matching predicate on an attribute of a relation, to be used in optimizing execution of a query directed to one or more attributes of said relation, comprising

providing a data storage structure storing character statistics on said attribute, comprising a first structure storing, for each of a plurality of character positions, frequently occurring characters in that character position, and statistics for each frequently occurring character,

retrieving said statistics in response to said pattern matching predicate based upon the character positions of characters in said pattern matching predicate, and

generating said statistic based upon said retrieved character statistics.
2. The method of claim 1 wherein said data storage structure further stores a count of a number of occurrences of a frequently occurring character.
3. The method of claim 1 wherein said data storage structure further comprises a second structure storing frequently occurring characters that are subsequent to the frequently occurring characters stored in said first structure.

4. The method of claim 3 wherein said second data storage structure further stores a probability of occurrence of a frequently occurring subsequent character.

5. The method of claim 1 wherein generating said statistic based upon said retrieved character statistics comprises estimating a statistic for a desired character and desired character position, in the event said statistic is not stored in said first structure.

6. The method of claim 5 wherein said estimating comprises accumulating statistics for said desired character position, for characters other than said desired character, and calculating average frequency of occurrence of characters that do not have statistics stored in said first structure.

7. The method of claim 1 wherein said first structure is a table.

8. The method of claim 1 wherein said first structure comprises a linked list.

9. The method of claim 3 wherein said second structure comprises a table.

10. The method of claim 3 wherein said second structure comprises a linked list.

11. A computer system implementing a relational database system and generating a statistic for a pattern matching predicate on an attribute of a relation of said relational database, to be used in optimizing execution of a query directed to one or more attributes of said relation, comprising

storage for said relational database, including a relation having a plurality of tuples including values for a plurality of attributes, and a data storage structure storing character statistics on an attribute, comprising a first structure storing, for each of a plurality of character positions, frequently occurring characters in that character position, and statistics for each frequently occurring character, and

computing circuitry performing query optimization and query execution upon said relational database, said query optimization including generating said statistic for an attribute of said relation by retrieving said statistics in response to said pattern matching predicate based upon the character positions of characters in said pattern matching predicate, and generating said statistic based upon said retrieved character statistics.

12. The computer system of claim 11 wherein said data storage structure further stores a count of a number of occurrences of a frequently occurring character.

13. The computer system of claim 11 wherein said data storage structure further comprises a second structure storing frequently occurring characters that are subsequent to the frequently occurring characters stored in said first structure.

14. The computer system of claim 13 wherein said second data storage structure further stores a probability of occurrence of a frequently occurring subsequent character.

15. The computer system of claim 11 wherein said computing circuitry generates said statistic based upon said retrieved character statistics by estimating a statistic for a desired character and desired character position, in the event said statistic is not stored in said first structure.

16. The computer system of claim 15 wherein said computer circuitry performs said estimating by accumulating statistics for said desired character position, for characters other than said desired character, and calculating average frequency of occurrence of characters that do not have statistics stored in said first structure.

17. The computer system of claim 11 wherein said first structure is a table.

18. The computer system of claim 11 wherein said first structure comprises a linked list.

19. The computer system of claim 13 wherein said second structure comprises a table.

20. The computer system of claim 13 wherein said second structure comprises a linked list.

21. A program product for implementing a relational database system and generating a statistic for a pattern matching predicate on an attribute of a relation of said relational database, to be used in optimizing execution of a query directed to one or more attributes of said relation, comprising

a relational database, including a relation having a plurality of tuples including values for a plurality of attributes,

a data storage structure storing character statistics on an attribute, comprising a first structure storing, for each of a plurality of character positions, frequently occurring characters in that character position, and statistics for each frequently occurring character, and

relational database software performing query optimization and query execution upon said relational database, said query optimization including generating a statistic for an attribute of said relation by retrieving said statistics in response to said pattern matching predicate based upon the character positions of characters in said pattern matching predicate, and generating said statistic based upon said retrieved character statistics, and a signal bearing media holding said relational database and relational database software.

22. The program product of claim 21 wherein the signal bearing media comprises transmission media.

23. The program product of claim 21 wherein the signal bearing media comprises recordable media.